TESTS

BY FRANK N. FREEMAN
University of Chicago

The war, and the large measure of success which was attained with the mental tests of soldiers, has stimulated an enormous amount of activity in the organization of new tests,—particularly group tests,—the further standardization or revision of old tests and the application of tests to the problems of education and of vocational guidance. From a general skepticism in regard to the possibility of making practical use of tests, popular opinion is swinging toward an undue faith in them; and it is to be feared that, unless psythologists are able to generate an attitude of caution, a reaction will follow the disappointment of false hopes.

Theory and Technique

In the midst of the production of new tests or their application to new problems there is some discussion of technique. This is desirable and we need more of it.

The theory which underlies the type of testing which is being so widely used—testing to discover a candidate's fitness for a job, particularly a complex one—is discussed by Thorndike (59). He points out that variations in the degree or amount of a trait which is requisite for a job may not correspond with proportional variations all along the line in the degree of efficiency in the job. The rise in efficiency may follow any one of a number of curves. Furthermore the individual traits in a complex ability may need to be weighted. But an element may contribute to two or more traits, and in this case it should not be weighted more than once. Finally, the value of traits which are interdependent may need to be multiplied rather than added.

The technique of examining the "efficiency" of a group test is illustrated by Pressey (44). The tests are tried out by giving them to three groups of children already widely differentiated by their social reactions—children in a feeble-minded institution, in a specially advanced class and the ordinary children of a public school. On this basis the whole scale is judged reliable and a shorter scale is selected.

The meaning and validity of the intelligence quotient are discussed by Mateer (29) and by Evans and Castle (10). Miss Mateer has followed 15 institutional cases who were seven years

old and had an I.Q. of .93 to .99 at the first test. Some of these have improved and give promise of becoming normal and others have actually retrograded. The significance of these facts seems to be that neuropathic conditions may produce irregularities in the development of intelligence. Evans and Castle (10) made a horizontal comparison between the kinds of achievement in different kinds of tests of 34 individuals, 15 of whom tested at age (by Yerkes-Bridges scale) and 19 above or below, mostly below. The chronologically older children were in general inferior in these other tests, particularly so in opposites, directions, memory for forms and letter squares. In five simpler tests the difference was slight. The authors conclude that the more complex tests are tests of brightness and the simpler ones tests of maturity. It would be more precise to say that the simpler tests (and also the Yerkes-Bridges) are more largely tests of maturity than the complex.

Methods of rendering tests free of error from coaching or from general or special practice are discussed by Thorndike (58). Among the devices of providing alternative forms, divising non-coachable tests, introducing confusion questions and comparing individuals' records in coachable and non-coachable parts of tests he prefers the first. He recommends fore-exercise and the use of fairly long tests.

The technique of presentation in the Knox Cube Test was studied by Rachofsky (47), who found that errors were fewer with slower presentation up to 2 sec. per tap.

Maxfield (30) gives formulæ by which one may calculate the number of children who may be expected to attain various mental ages from the application of tests for which we know the number at each age level and the frequency with which each is passed by unselected children. Kohs (21) has devised a convenient slide rule for calculating I. Q's.

Studies of Old Tests

A rather elaborate study of a number of tests which had previously been worked out as group tests by Pyle is presented by Pintner (35) in *The Mental Survey*. In view of the many ingenious group tests which are being prepared these earlier tests are undoubtedly superseded, but the methods of organizing the results which Pintner elaborates are very useful.

A comparison of the Binet and Yerkes-Bridges scales with 50 high-grade defectives is reported by Lewis (23) but his results are

not very conclusive. Sunne (51) compared the diagnostic value of the individual tests of the Yerkes-Bridges scale with 550 children, two groups of white and two of colored. Large differences in value were found but they varied with the different groups.

The vocabulary test has been studied by Terman (55) and shown to have a rather surprisingly high diagnostic value (in terms of the Stanford revision), the correlation being .91 for children and .81 for adults, and to be little affected by chronological age as compared with mental age. Foreign speaking children test to age above 12 years of age. Porteus (40) finds that his maze test correlates about .7 with the Binet scores on over 600 children. It appears to measure temperament to some extent and he reports that delinquents are particularly deficient in it. Other details are given.

Studies of memory tests are reported by Gates (12) and by King and Homan (20). Gates presents correlations between immediate and delayed recall, between the tests and teachers' estimates and between memory of sense and nonsense material, among others. King and Homan compare correlations of historical, descriptive and narration material and material of different length.

Moon (31) gives a descriptive summary of age scales.

New Tests

The army tests are referred to in an anonymous article in *Science* (1). The most distinctive feature of the main scale, Scale Alpha, is that it is a group test. The army tests represent a coöperative undertaking. The directing head in the organization of these tests was Yerkes, but many psychologists coöperated with him both officially and unofficially. The point-scale organization has direct relation to Yerkes's previous study with Bridges and Hardwick and to the further refinement of the point-scale method by Otis. The content of the tests is derived from many sources, particularly of Scale Alpha, which includes a very simple directions test, arithmetic reasoning, checking best reasons, opposites, completion of number series, analogies, and information. Scale Beta for illiterates, a non-verbal test, contains a number which had been organized by Thorndike. The Stanford Revision, the Yerkes-Bridges test, the Pintner-Paterson performance scale, the Stenquist test, etc., were used for individual examination. In Scale Alpha each test consists of from ten to forty parts graded in difficulty. The mechanics of the response and of scoring are very simple.

The group tests described by Thorndike (57) may be performed

without the use of language. They include the processes of digit-symbol substitution, completing pictures, picture analogies, easy computation, dividing a surface to correspond to separate figures which may be made to fit it, completing rhythmic series of forms and mixed spatial relations. Ten alternate forms are given and little emphasis is put on speed. Various correlations are given in the report.

A point scale very similar to the army Scale Alpha is published by Otis (32). This is a group scale of ten tests, each composed of a number of similar units. There are two parallel forms and the test is designed for the upper grades. The responses required are very simple and the scoring mechanical. A group point scale designed for the high-school level consisting of ten tests, each of a number of units, is described by S. L. and L. W. Pressey (43). There are several novel tests in the series, as logical selection, moral classification and a new form of practical information test. The authors have also another series of tests which require only the response of crossing out one element, and a primer scale which consists of four tests put in pictorial form. Thurstone at Carnegie Institute of Technology has published a test for high-school graduates in which the arrangement of the units of the different tests is spiral. The easy units of all the tests are placed first, then the next harder units, etc. This makes it unnecessary to time each test separately. Thurstone also publishes a clerical examination. Thorndike's intelligence examination for high school graduates is widely known. Its distinctive features are its length and large number of parts designed to overcome the influence of chance errors, the many parallel forms which are provided and the large variety of kinds of tests which include, beside such as are in the army scales, the elaborate information and reading tests.

A very interesting scale of reasoning problems is given by Burt (5). There are 50 problems, many of them of the syllogistic type, arranged by ages, about six to each age. The problems are located according to median performance. A short scale of 17 problems is also designated. Herring's tests (16) also measure reasoning but break it up into various constituent processes, such as the judgment of the value of problems, of the feasibility of solutions, of the accuracy of definitions, of the clarity of statements, etc. They are designed to comprise scientific thinking. Results are not yet presented.

An elaborate study of age progress and of various correlations

in the case of a series of individual tests not organized into a scale is reported by Bickersteth (3). There are two motor tests (one new), tests of discriminative selection (new), three memory tests, the spot pattern test, a dotting test, a test of divided attention, a completion and an analogies test.

Mateer (28) presents an elaborate study of the conditioned reflex adapted to use with young children and used in conjunction with other tests. She believes that the rate of unlearning is a particularly valuable diagnostic indication, better for some cases than any other test.

A number of individual tests have been standardized. This includes in most cases selection and arrangement of new material, careful determination of the technique of presentation and scoring, and calculation of age norms. The following are included in the bibliography: a picture completion test by Shaw (50); another picture completion test by Lindley (24); a drawing completion test and a revised directions test by Pintner and Toops (38, 39); an opposites test in point-scale form by Greene (14); a false definition test by Gerlach (13); a vocabulary test by Brandenburg (4); and an accuracy of movement test by Beeley (2).

A test based on the judgment of character by associates is reported by Robson (48). Pintner (36) throws some light on judgment by a study of physiognomy by having the intelligence of a number of children judged from their photographs. The median correlation from several groups of judges was about .10.

Applications of Tests

The group of reports on this topic must be treated very briefly. A number of studies have been made dealing with the use of tests to diagnose the ability required to do the work of the school or the college. A valuable summary of a large amount of study of this problem, made under the direction of Terman, is presented in his *Intelligence of School Children* (52). This book gives an impressive array of facts which indicate that "intelligence," or the sort of thing which is measured by tests, is a very large factor in school success. The studies of Cuneo and Terman (9) and of Proctor (45) are among those summarized by Terman. Miss Race (46) reports the rapid progress of a special class, selected by tests. Toops and Pintner (60) report that the distribution of the test ratings of unemployed men corresponds to the distribution of the grades at which they left school. This, however, does not establish a

correlation. Studies on the college level are reported by Haggerty (15), Hill (17), James (18) and King and McCrory (20).

A second type of application deals with the use of tests to diagnose fitness for a vocation. A considerable measure of success in this type of endeavor is reported by Link (25, 26). Pintner and Toops (37) show that lack of success in general in maintaining a permanent economic independence goes with mental deficiency. Flanders (11) however, found practically no correlation between the I.Q.'s of express clerks and their degree of success. Finally, Luckey (27) found that feeble-minded childrens' improvement in the form board correlated well with their rating in industrial improvement.

A comparatively new type of work is represented in the mental survey of whole communities [Pintner (35), Paterson (33), Pressey (42)]. The practical value of this work is still somewhat an open question.

The hereditary and environmental factors in ability constitute the last type of problem. Kornhauser (22) shows that economic position is related to ability in school work. Pintner (34) shows that siblings are slightly more similar than chance groups. L. W. Pressey (41) studied sex differences and found girls slightly superior, but not in all tests, and boys more variable.

References

1. Anon: The Measurement and Utilization of Brain Power in the Army. *Science*, 1919, **49**, 221–226; 251–259.
2. Beeley, A. L. *An Experimental Study in Left-Handedness.* (Supplementary Monographs, Dept. of Educ., University of Chicago, 1918.) Pp. 74.
3. Bickersteth, M. E. The Application of Mental Tests to Children of Various Ages. *Brit. J. of Psychol.*, 1918, **9**, 23–73.
4. Brandenburg, G. C. Psychological Aspects of Language. *J. of Educ. Psychol.*, 1918, **9**, 313–332.
5. Burt, C. Development of Reasoning in Children. *J. of Exper. Ped.*, 1919, **5**, 68–77.
6. Carlisle, C. L. Performance Norms for Thirteen Tests and Mental Examinations. *J. of Educ. Psychol.*, 1918, **9**, 518–523.
7. Cleveland, E. Measuring Failure. *School & Soc.*, 1918, **8**, 205–209, 236–239.
8. Cole, L. W. Mental Age and School Entrance. *School & Soc.*, 1918, **8**, 418–419.
9. Cuneo, I., & Terman, L. M. Stanford-Binet Tests of 112 Kindergarten Children and 77 Repeated Tests. *Ped. Sem.*, 1918, **25**, 414–428.
10. Evans, J. E. & Castle, M. The Relation of Mental Age to Chronological Age as determined by Certain Individual and Group Tests. *J. of Appl. Psychol.*, 1918, **2**, 308–323.
11. Flanders, J. K. Mental Tests of a Group of Employed Men showing Correlations with Estimate Furnished by Employer. *J. of Appl. Psychol.*, 1918, **2**, 197–206.

12. GATES, A. I. Correlations of Immediate and Delayed Recall. *J. of Educ. Psychol.*, 1918, **9**, 489–496
13. GERLACH, F. M. *Vocabulary Studies.* (Colorado Studies in Educ. & Psychol., 1907.) Pp. 123.
14. GREENE, H. A. A Standardization of Certain Opposites Tests. *J. of Educ. Psychol.*, 1918, **9**, 559–566.
15. HAGGERTY, M. E. Tests of Applicants for Admission to University of Minnesota Medical School. *J. of Educ Psychol.*, 1918, **9**, 278–286.
16. HERRING, J. P Measurements of some Abilities in Scientific Thinking. *J. of Educ. Psychol.*, 1918, **9**, 535–558.
17. HILL, D. S. Results of Intelligence Tests at the University of Illinois. *School & Soc.*, 1919, **9**, 542–545.
18. JAMES, B. B. Correlations of Mental Tests and Scholarship. *School & Soc.*, 1918, **7**, 238–239.
19. KING, I. & HOMAN. Logical Memory and School Grades *J. of Educ. Psychol.*, 1918, **9**, 262–269.
20. KING, I. & MCCRORY, J. Freshman Tests at State University of Iowa. *J. of Educ. Psychol*, 1918, **9**, 32–46.
21. KOHS, S. C. The Reed I Q. Slide Rule. *School & Soc.*, 1918, **8**, 749–750.
22. KORNHAUSER, A. W. The Economic Standing of Parents and the Intelligence of Children. *J. of Educ. Psychol.*, 1918, **9**, 159–164.
23. LEWIS, E. O. The Binet and Point-Scale Methods of Testing Intelligence. *J. of Exper. Ped.*, 1918, **4**, 198–202.
24. LINDLEY, M. A Picture Test. *Training Sch Bull*, 1918, **14**, 156–161.
25. LINK, H. C. An Experiment in Employment Psychology. *Psychol. Rev.*, 1918, **25**, 116–127.
26. LINK, H. C. *Employment Psychology.* New York: Macmillan, 1919 Pp. xii + 440.
27. LUCKEY, B. M. Correlation between Form-board Improvement and Ability to Improve Industrially. *Training Sch. Bull.*, 1918, **14**, 140–142.
28. MATEER, F. *Child Behavior.* Boston: Badger, 1918.
29. MATEER, F. The Diagnostic Fallibility of Intelligence Ratios. *Ped. Sem.*, 1918, **25**, 369–392.
30. MAXFIELD, F. N. Some Mathematical Aspects of the Binet-Simon Tests. *J. of Educ. Psychol.*, 1918, **9**, 1–12
31. MOON, R. C. Age Scale Methods of Measuring Intelligence. *J. of Exper. Ped.*, 1919, **5**, 78–97.
32. OTIS, A. S. An Absolute Point Scale for the Group Measurements of Intelligence. *J. of Exper. Psychol.*, 1918, **9**, 239–261, 333–348.
33. PATERSON, D. G. A Mental Survey of the School Population of a Kansas Town. *School & Soc.*, 1918, **7**, 84–89.
34. PINTNER, R. The Mental Indices of Siblings. *Psychol. Rev.*, 1918, **25**, 252–255.
35. PINTNER, R. *The Mental Survey.* New York: Appleton, 1918. Pp. 116.
36. PINTNER, R. Intelligence as Estimated from Photographs. *Psychol. Rev.*, 1918, **25**, 286–296.
37. PINTNER, R., & TOOPS, H. A. Mental Tests of Unemployed Men. *J. of Appl. Psychol.*, 1918, **2**, 15–25.
38. PINTNER, R. & TOOPS, H. A. A Drawing Completion Test. *J. of Appl. Psychol.*, 1918, **2**, 164–173.

39. PINTNER, R., & TOOPS, H. A. A Revised Directions Test. *J. of Educ. Psychol.*, 1918, **9**, 123–142.
40. PORTEUS, S. D. The Measurement of Intelligence· Six Hundred and Fifty-three Children Examined by the Binet and Porteus Tests. *J. of Educ. Psychol.*, 1918, **9**, 13–31.
41. PRESSEY, L. W. Sex Differences Shown by 2,544 School Children on a Group Scale of Intelligence, with Special Reference to Variability. *J. of Appl. Psychol.*, 1918, **2**, 323–340.
42. PRESSEY, S. L. Comparison of Two Cities and their School Systems by Means of a Group Scale of Intelligence. *Educ. Admin. & Supervision*, 1919, **5**, 53–62.
43. PRESSEY, S. L., & PRESSEY, L. W. A Group Point Scale for Measuring General Intelligence, with First Results from 1,100 School Children. *J. of Appl. Psychol.*, 1918, **2**, 250–269.
44. PRESSEY, S. L., & PRESSEY, L. W. The Practical "Efficiency" of a Group Scale of Intelligence. *J. of Appl Psychol*, 1919, **3**, 68–80.
45. PROCTOR, W. M. The Use of Intelligence Tests in the Educational Guidance of High-School Pupils. *School & Soc*, 1918, **8**, 473–478, 502–509.
46. RACE, H. V. A Study of a Class of Superior Children. *J. of Educ. Psychol.*, 1918, **9**, 91–98.
47. RACHOFSKY, L. M. Speed of Presentation and Ease of Recall in the Knox Cube Test. *Psychol. Bull.*, 1918, **15**, 61–64.
48. ROBSON, M. M An Experimental Study of Character. *J. of Educ. Psychol.*, 1918. **9**, 514–516.
49. ROSANOFF, A. J., MARTIN, H. E., & ROSANOFF, I. R. A Higher Scale of Mental Measurement and Its Application to Cases of Insanity. *Psychol. Monog.*, 1918, **25**, 1–113.
50. SHAW, E. A. A New Picture Completion Test. *J. of Appl. Psychol.*, 1918, **2**, 355–365.
51. SUNNE, D. The Diagnostic Value of the Yerkes Point Scale. *School & Soc.*, 1918, **8**, 387–389.
52. TERMAN, L. M. The Intelligence of School Children. Boston: Houghton Mifflin, 1919.
53. TERMAN, L. M. The Use of Intelligence Tests in the Army. PSYCHOL. BULL., 1918, **15**, 177–186.
54. TERMAN, L. M. *The Stanford Revision and Extension of the Binet Simon Scale for Measuring Intelligence* Baltimore: Warwick & York, 1918. Pp. 179.
55. TERMAN, L M. The Vocabulary Test as a Measure of Intelligence. *J. of Educ. Psychol.*, 1918, **9**, 452–466.
56. TERMAN, L. M. & CHAMBERLAIN, M. B. Twenty-three Serial Tests of Intelligence and their Intercorrelations. *J. of Appl. Psychol.*, 1918, **2**, 341–354.
57. THORNDIKE, E. L. A Standardized Group Examination of Intelligence Independent of Language. *J. of Appl. Psychol.*, 1919, **3**, 13–32.
58. THORNDIKE, E. L. Tests of Intelligence; Reliability, Significance, Susceptibility to Special Training and Adaptation to the General Nature of the Task. *School & Soc.*, 1919, **9**, 189–195.
59. THORNDIKE, E. L. Fundamental Theorems in Judging Men. *J. of Appl. Psych.*, 1918, **2**, 67–76.
60. TOOPS, H. A., & PINTNER, R. Mentality in its Relation to Elimination from Schools. *School & Soc.*, 1918, **7**, 507–510, 534–539.